From Radio Waves to Gamma Rays: A Spectral Analogy for Biomedical Data Science
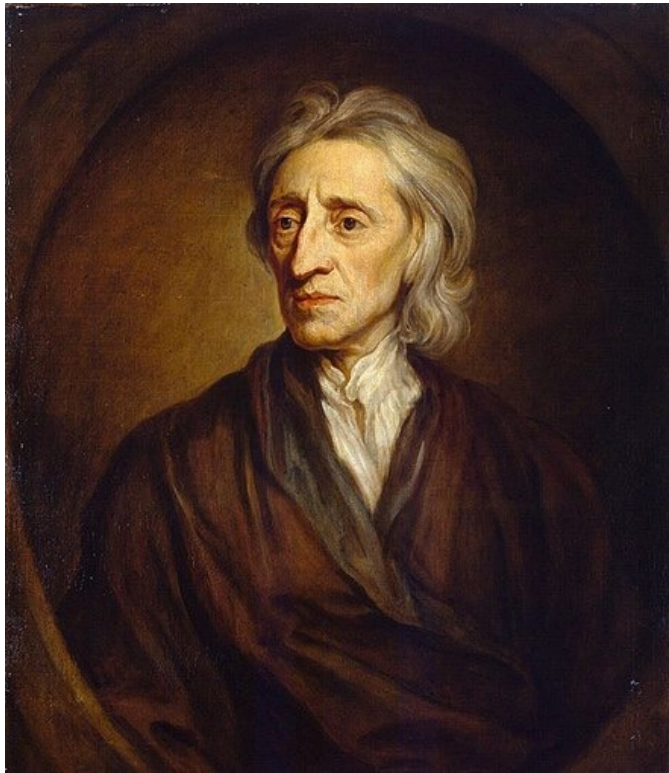
University of Pennsylvania May 23, 2019

Prelimninaries

John Locke on Tolerance

"For every church [discipline] is orthodox to itself; to others, erroneous or heretical."
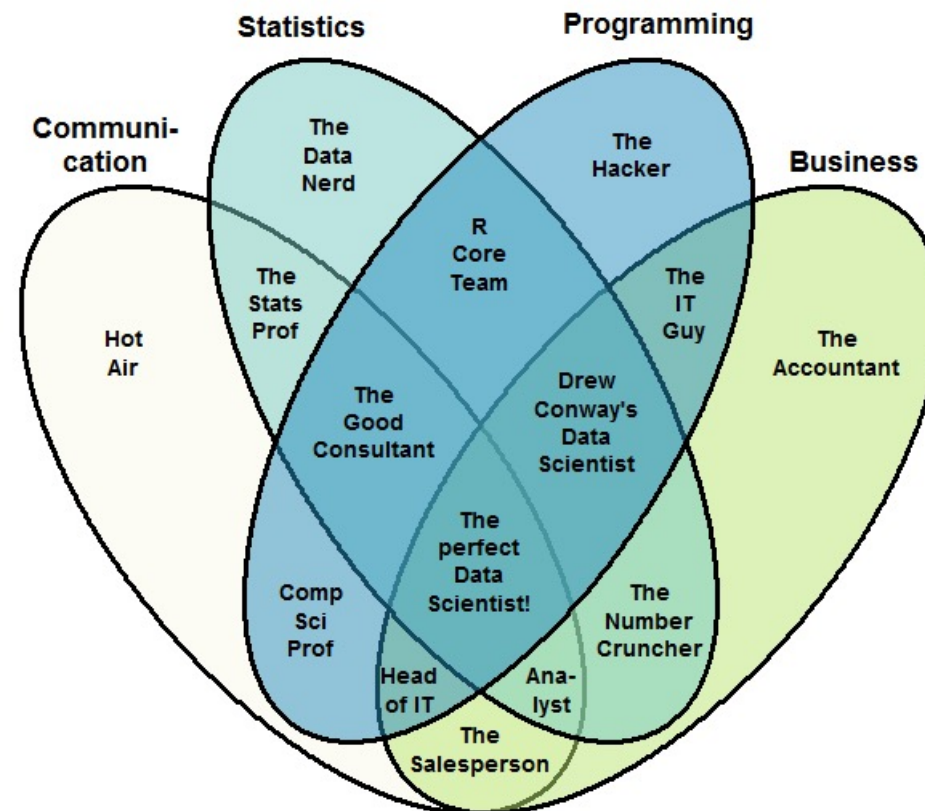
("A Letter concerning Toleration")

"Hymn to Him"

"Why can't a woman be more like a man?

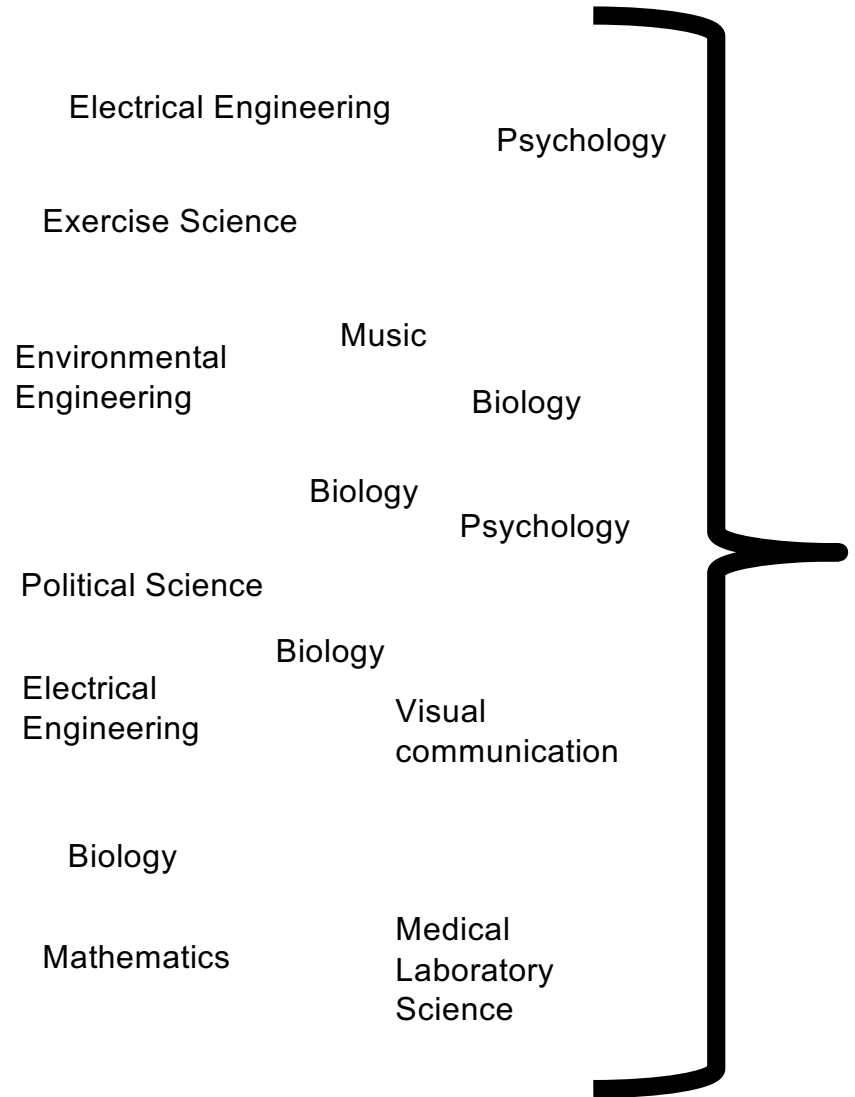Men are so decent, such regular chaps...

Why can't a woman be like me?"

What is Data Science? Who can be a Data Scientist

The Data Scientist Venn Diagram

• Wikipedia

Electrical Engineering

Psychology

Exercise Science

Music

Environmental Engineering

Biology

Biology

Psychology

Political Science

Biology

Electrical Engineering

Visual communication

Biology

Mathematics

Medical Laboratory Science

Biomedical Data Scientist

Is this Diversity Lamentable?

"[Kepler] is only one of many examples of the falsity of the idea that success in scientific research demands an exclusive absorption in one narrow line of study. Novel ideas are more apt to spring from an unusual assortment of knowledge—not necessarily from vast knowledge, but from a thorough conception of the methods and ideas of distinct lines of thought."
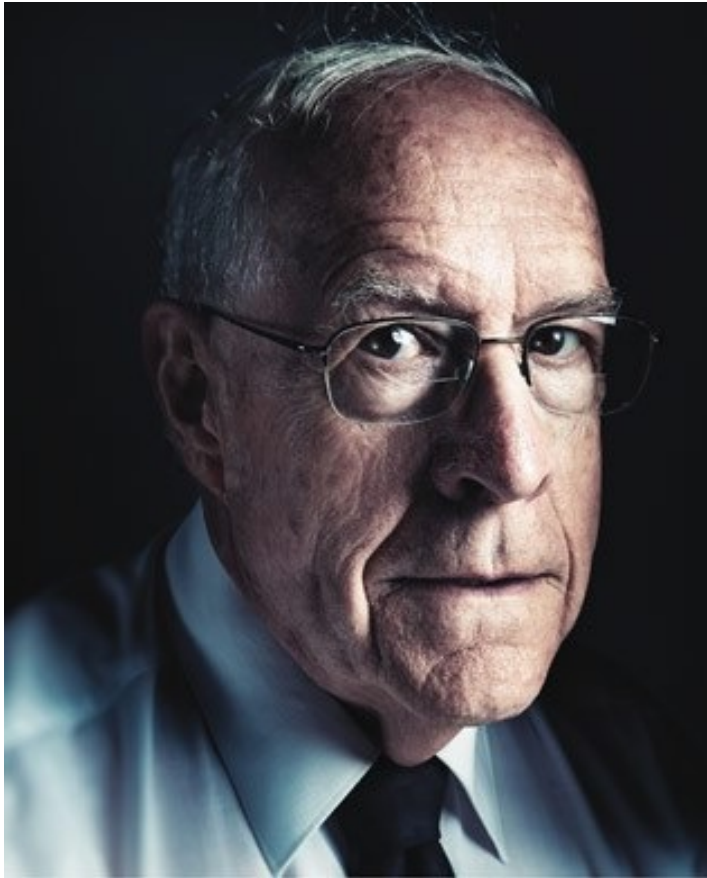
Alfred North Whitehead. An Introduction to Mathematics (p. 112).

I'm a Fred Brooks Disciple

Fred Brooks on Names and Naming Errors

www.wired.com

"What's in a name? Much. Our self-misnaming hastens various unhappy trends…. **[W]e shall be respected for our accomplishments, not our titles.**" ("Computer Scientist as Toolsmith")

What Should AI Do?

	Human-Based	Ideal Rationality
Reasoning-Based:	Systems that think like humans.	Systems that think rationally.
Behavior-Based:	Systems that act like humans.	Systems that act rationally.

Four Possible Goals for AI According to AIMA

Bringsjord, Selmer and Govindarajulu, Naveen Sundar, "Artificial Intelligence", *The Stanford Encyclopedia of Philosophy* (Fall 2018 Edition), Edward N. Zalta (ed.),
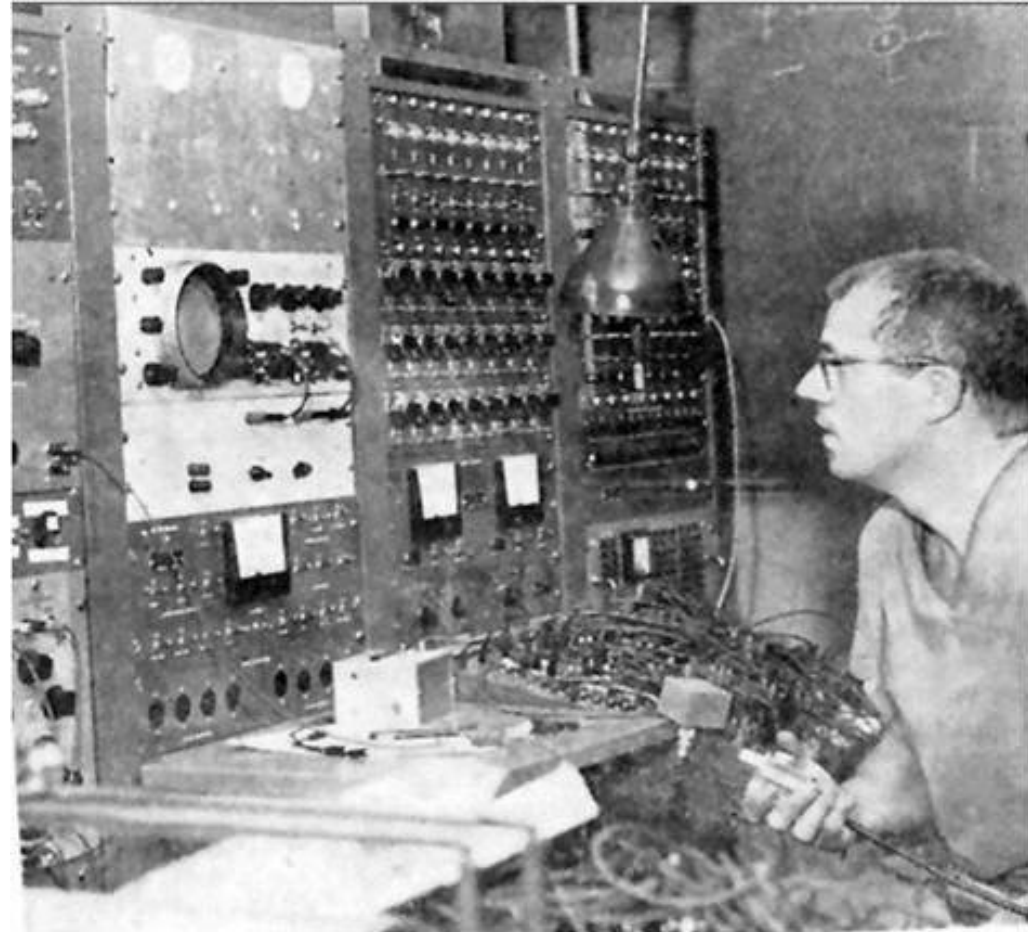
And What of Biomedical Data Science?

- Use data to facilitate creating human and artificial systems that facilitate clinicians and patients acting rationally

What Should Biomedical Informatics Do?

- Create human and artificial systems that facilitate clinicians and patients acting rationally

Have the Tasks Changed

Homer Warner with his
first computer circa
1960

- https://www.nytimes.com/2012/12/11/us/homer-r-warner-a-pioneer-of-using-computers-in-patient-care-dies-at-90.html

BIOMEDICAL INFORMATICS

9 The thing that hath been, it is that which shall be; and that which is done is that which shall be done: ***and there is no new thing under the sun.***
10 Is there any thing whereof it may be said, See, this is new? it hath been already of old time, which was before us. Ecclesiastes

"Fifty Years of Data Science"

One doesn't need to look far to see click-bait capitalizing on the befuddlement about this new state of affairs:

- *Why Do We Need Data Science When We've Had Statistics for Centuries?*
  Irving Wladawsky-Berger
  Wall Street Journal, CIO report, May 2, 2014

- *Data Science* **is** *statistics.*

  ```
  When physicists do mathematics, they don't say they're doing number science.  They're doing
  math.  If you're analyzing data, you're doing statistics.  You can call it data science
  or informatics or analytics or whatever, but it's still statistics.  ...  You may not like
  what some statisticians do.  You may feel they don't share your values.  They may embarrass
  you.  But that shouldn't lead us to abandon the term ``statistics''.
  ```

  Karl Broman, Univ. Wisconsin[6]

On the other hand, we can find pointed comments about the (near-) irrelevance of statistics:

- *Data Science without statistics is possible, even desirable.*
  Vincent Granville, at the Data Science Central Blog[7]

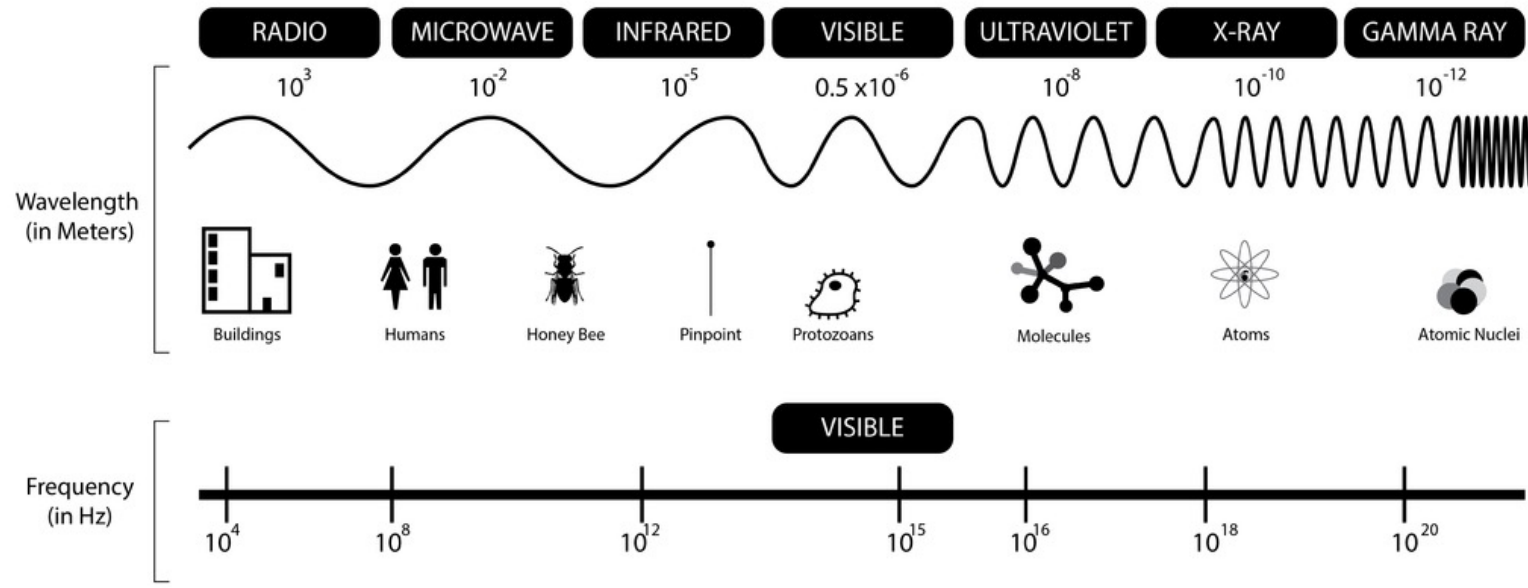- *Statistics is the least important part of data science.*
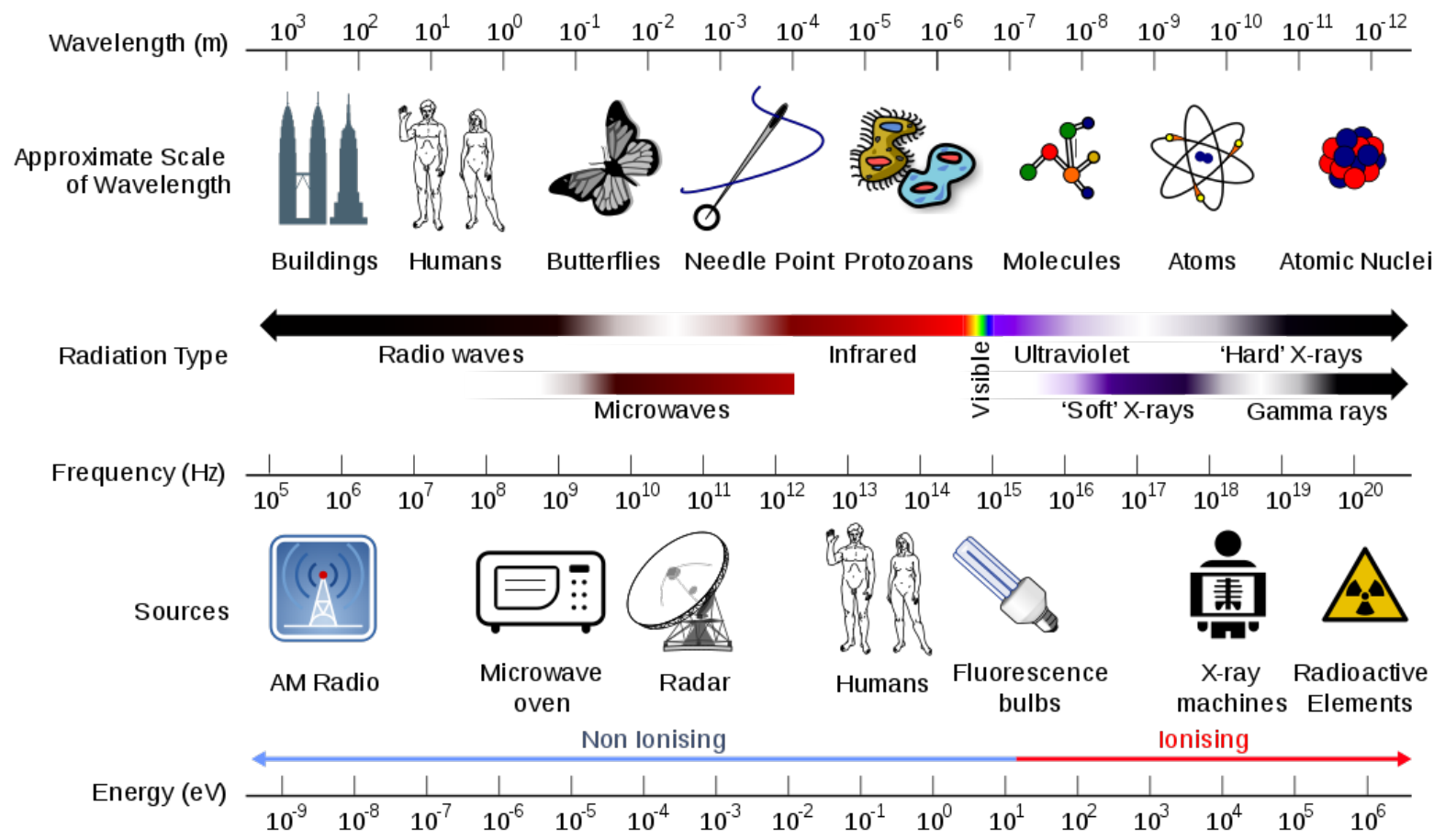 Andrew Gelman, Columbia University [8]

On To My Analogy

U HEALTH
UNIVERSITY OF UTAH

THE ELECTROMAGNETIC SPECTRUM

RADIO	MICROWAVE	INFRARED	VISIBLE	ULTRAVIOLET	X-RAY	GAMMA RAY
10^3	10^{-2}	10^{-5}	0.5×10^{-6}	10^{-8}	10^{-10}	10^{-12}

Wavelength (in Meters)

Buildings	Humans	Honey Bee	Pinpoint	Protozoans	Molecules	Atoms	Atomic Nuclei

VISIBLE

Frequency (in Hz)

10^4 10^8 10^{12} 10^{15} 10^{16} 10^{18} 10^{20}

The electromagnetic spectrum

Wavelength (m): 10^3 10^2 10^1 10^0 10^{-1} 10^{-2} 10^{-3} 10^{-4} 10^{-5} 10^{-6} 10^{-7} 10^{-8} 10^{-9} 10^{-10} 10^{-11} 10^{-12}

Approximate Scale of Wavelength: Buildings · Humans · Butterflies · Needle Point · Protozoans · Molecules · Atoms · Atomic Nuclei

Radiation Type: Radio waves · Microwaves · Infrared · Visible · Ultraviolet · 'Soft' X-rays · 'Hard' X-rays · Gamma rays

Frequency (Hz): 10^5 10^6 10^7 10^8 10^9 10^{10} 10^{11} 10^{12} 10^{13} 10^{14} 10^{15} 10^{16} 10^{17} 10^{18} 10^{19} 10^{20}

Sources: AM Radio · Microwave oven · Radar · Humans · Fluorescence bulbs · X-ray machines · Radioactive Elements

Non Ionising · Ionising

Energy (eV): 10^{-9} 10^{-8} 10^{-7} 10^{-6} 10^{-5} 10^{-4} 10^{-3} 10^{-2} 10^{-1} 10^0 10^1 10^2 10^3 10^4 10^5 10^6
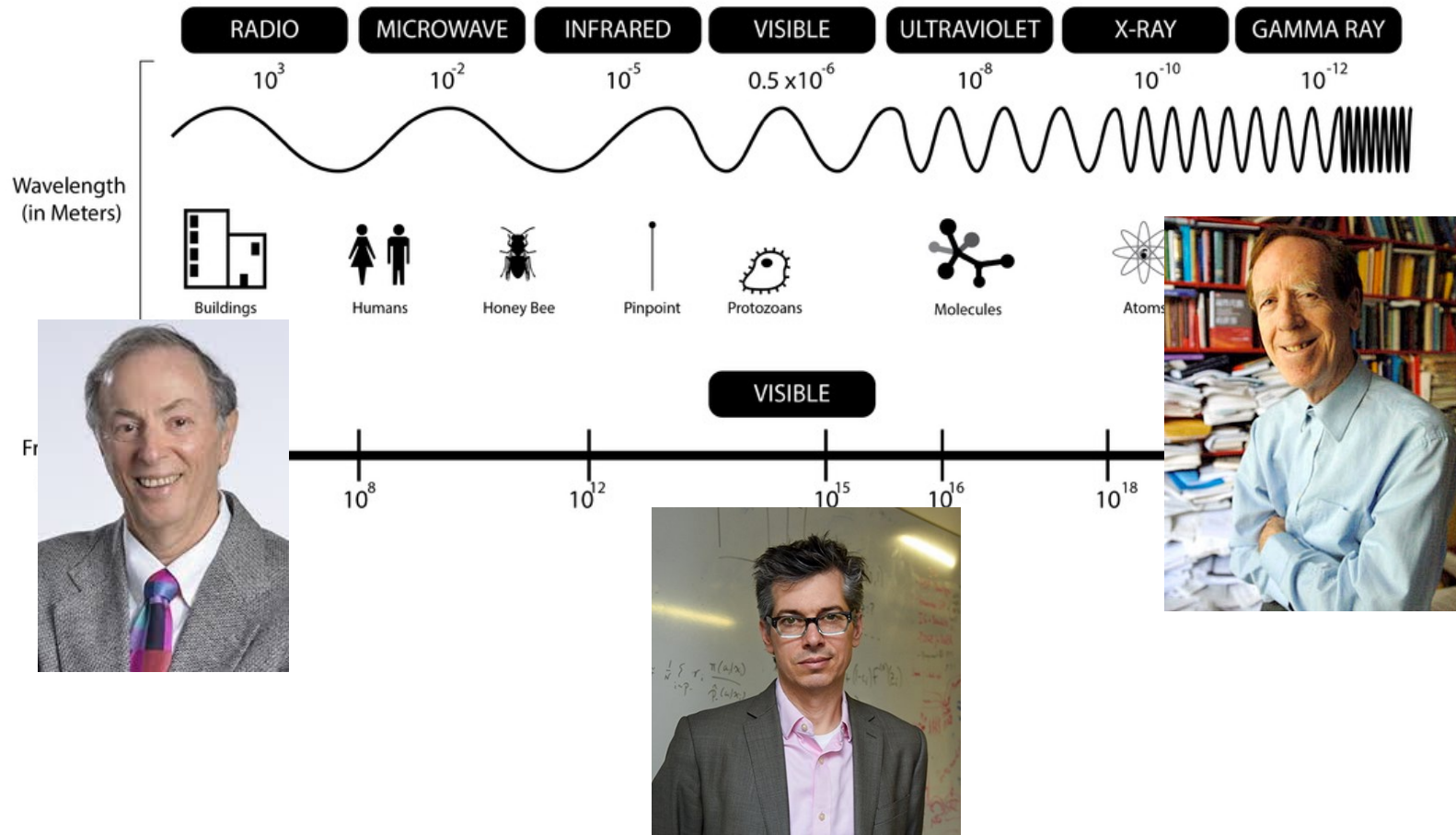
HEALTH
UNIVERSITY OF UTAH

Is Everything the Same Along the Spectrum?

- It is all electromagnetics!
- But…
  - Low Frequencies I might ignore wave properties
    - Kirchhoff's Current and Voltage Laws
  - At High Frequencies I might also ignore wave properties
    - Think in terms of photons
- Different **Skill Sets and Tools** to be productive
  - It is all Electromagnetics!

THE ELECTROMAGNETIC SPECTRUM

RADIO | MICROWAVE | INFRARED | VISIBLE | ULTRAVIOLET | X-RAY | GAMMA RAY

10^3 | 10^{-2} | 10^{-5} | 0.5×10^{-6} | 10^{-8} | 10^{-10} | 10^{-12}

Wavelength (in Meters)

Buildings | Humans | Honey Bee | Pinpoint | Protozoans | Molecules | Atoms

VISIBLE

Fr...

10^8 | 10^{12} | 10^{15} | 10^{16} | 10^{18}

Gilbert Strang on Data Science

- Linear Algebra
- Optimization
  - Stochastic gradient descent
- Probability
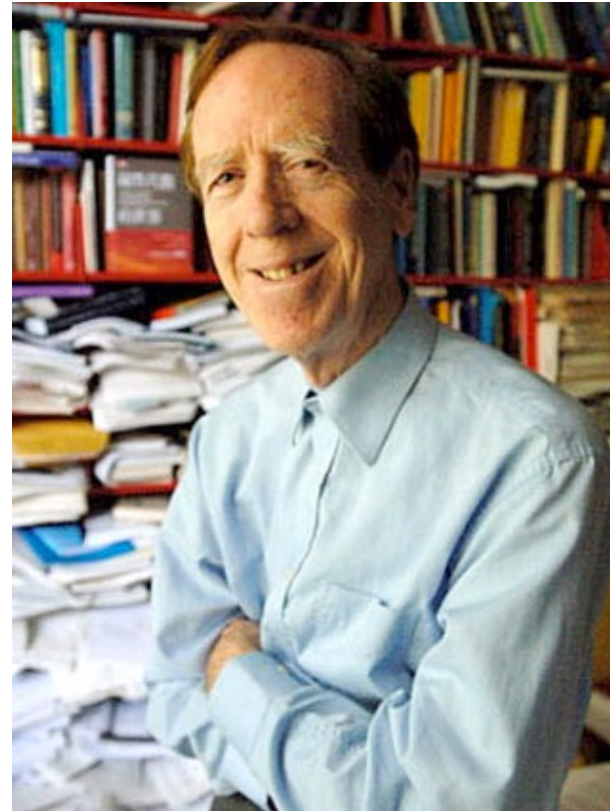
Chris Wiggins on Data Science

- What are people hired to do?
  - Query a SQL database
  - Analyze data with scikit-learn
  - Provide visualizations with Flask

Avi Seidmann on Data Science

- Keep Leaders from being bamboozled
  - TensorFlow for CEOs

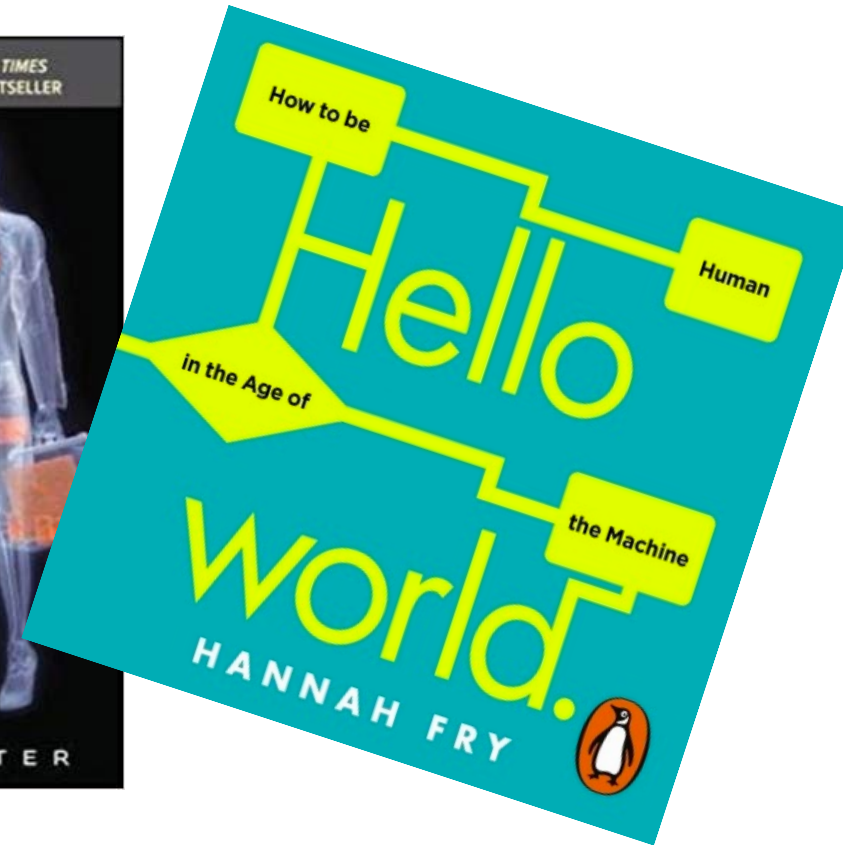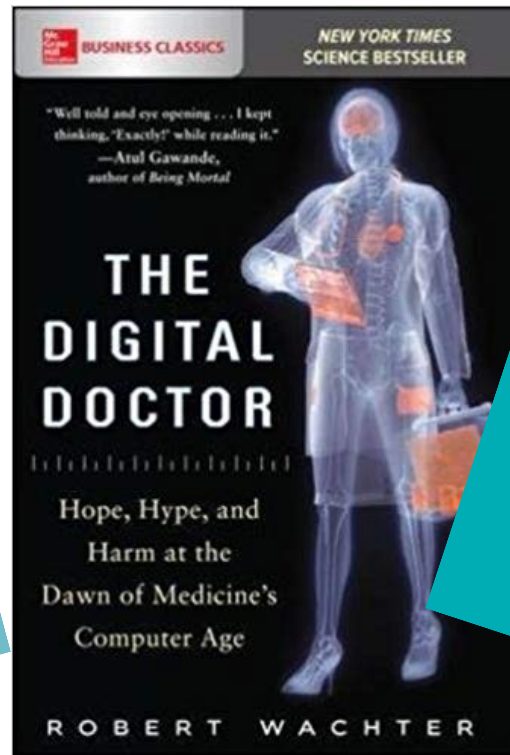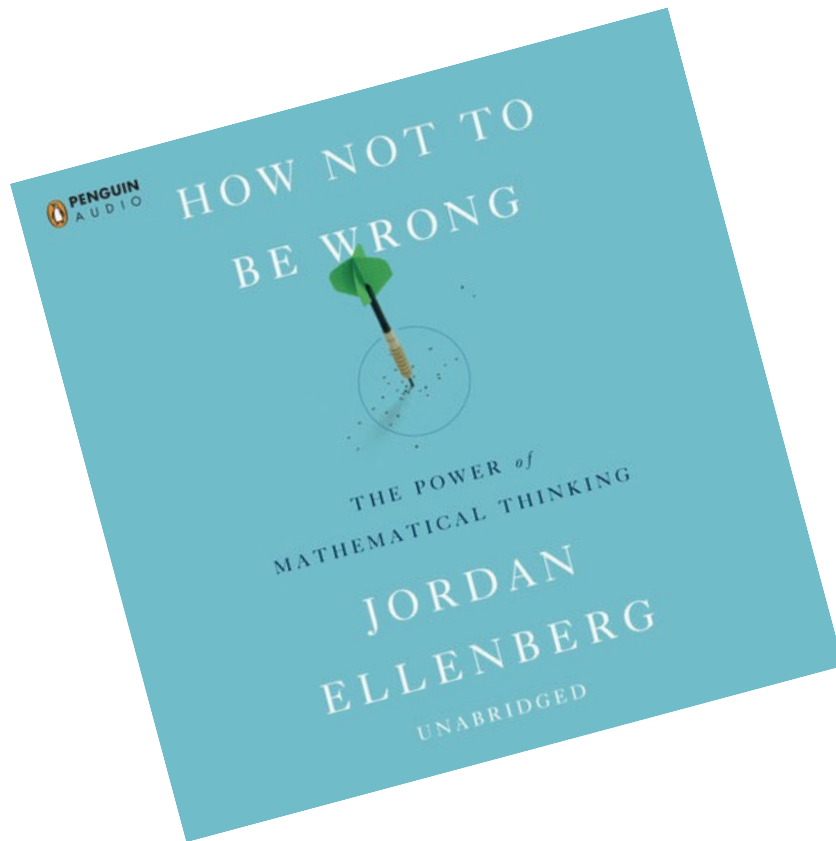# Is Biomedical Data Science Different?

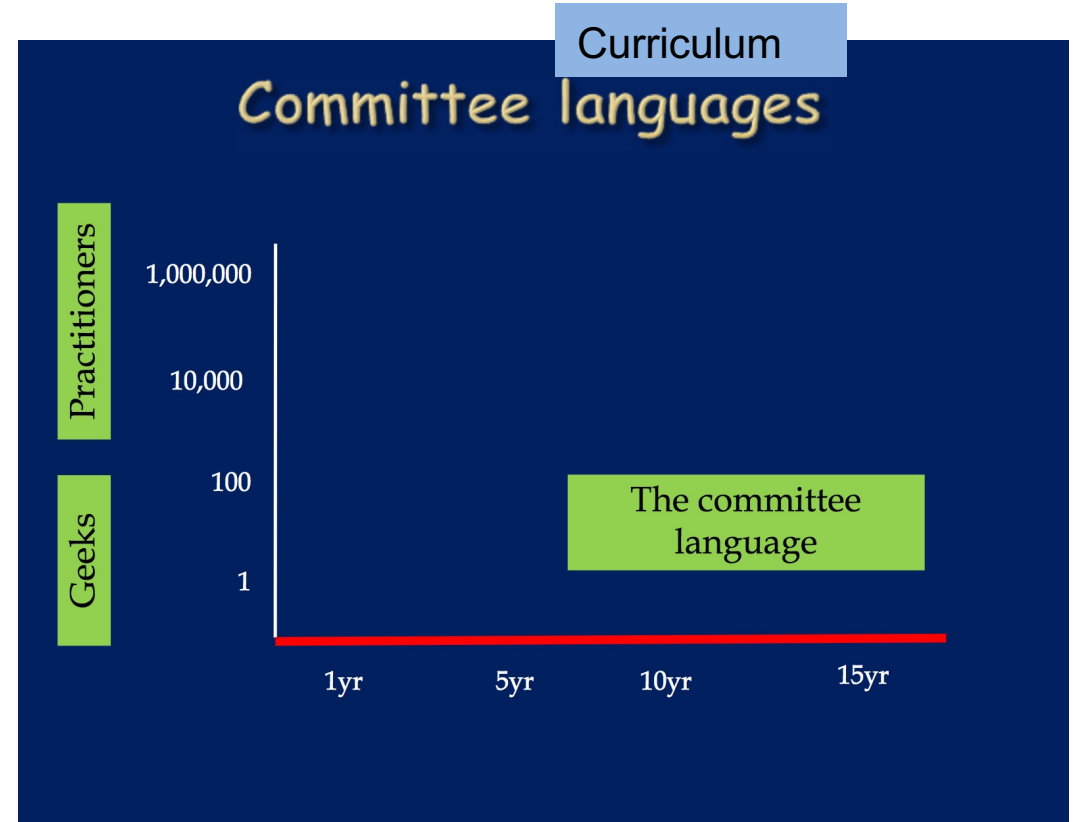# Teaching Novices Biomedical Data Science

- We're Diverse

- Task Driven

  - **Biomedical Objectives First**

- Integrated **Throughout** Curriculum

  - Methods not the ends

- Learn to Read then Write

Reading Biomedical Data Science

Designing a Biomedical Data Science Curriculum

- Modular
- Challenging
  - Supportive
- Broad Spectrum
- Evolving
- **Beware of the "Hymn to Him" syndrome**

Simon Peyton Jones

Regardless of the Season, the task remains the same